The Effects of AUGMENTED REALITY on HUMAN PERFORMANCE

The Bottom Line Up Front

Using Augmented Reality (AR) to provide work instructions to assembly technicians causes an instantaneous *human performance enhancement* (HPE).

- Direct build-time **speed increases by factor 3.5 times**
- **Standard deviation is less than *one third*** of doing the same task using traditional methods
- **Errors are reduced by 92%**.

For a wide array of assembly tasks, a high quality AR system can provide an *instantaneous* 70% reduction in the costs of value-added work, with *no risk* to your assembly and *no installation down-time* on your production line.

DSC has installed AR systems on many aircraft production lines. AR is a complicated subject and applies in different ways to different types of tasks. To the uninitiated, the application of AR seems daunting and risky. We often hear "I can see that AR is better than traditional methods to convey complex information – but how much better? I need a number." For quite some time we have struggled on how we might answer this question with sufficient fidelity without compromising any of our non-disclosure agreements.

We recently realized that we had been thinking about it all wrong. Because AR is commonly applied to rather complex assemblies, we had always envisioned that any fair comparison between AR and traditional assembly methods would, of necessity, need to be complex. Upon deeper reflection, we realized that is not true. The test only needs to be a very strictly controlled experiment where an identical task is performed by a large number of people in two ways (with and without AR), with accurate measurements of time and errors. *The use of AR is the only difference between the two trials, **so the difference in the measured values of time and errors represent the human performance enhancement caused by AR.*** *The critical factor being measured is **how** information flows from the engineer to the technician – not what is contained in that information.*

That last statement is key to understanding how to apply the results of this experiment to an actual production process where complex and critical assembly is taking place. The performance enhancement is coming from ***how*** the assembly instructions were transferred into the mind of the assembler – that is the only difference between the two trials. As it turns out, *how* information is transferred makes a HUGE difference in how quickly and comprehensively it is understood by the receiver of that information. To achieve these remarkable gains in productivity – how you transfer the work instructions from the engineer to the technician really matters. The consequences are enormous. The best part of all – the workers are not working any harder than before to produce over three times more product. All of the gains are purely from improved efficiency.

Backing Up the Claims

The claims above seem phenomenal; way too good to possibly be true. We are fully aware that these claims seem absurd and unrealistic. This paper will start with reporting the statistical data that forms the basis of these claims. As we progress deeper into the paper we will get deeper "into the weeds" on how the experiment was conducted. It is our hope that you will be able to make your own assessment on how the results of this very controlled and measured experiment may apply to your situation.

The implications for such extreme HPE can be staggering. A relatively minor implementation can result in millions of dollars in savings. Technicians can freely move from station to station and perform work in a station that is completely new to them and produce error-free assemblies in about the same time as a person that works in that cell on a daily basis (which is less than half the time it takes that experienced person today). *A 100% return on investment is typically measured in weeks (15-30)*. More on that can be read in the companion paper to this one entitled "*Return on Investment, The Application of Augmented Reality on Complex Assembly*".

Here are the main statistics:

Number of Participants **85**

Average Build Time **Delta 0.29x**

- by Drawing 80.47 sec
- by AR 23.67 sec

Standard Deviation **Sigma 0.32x**

- by Drawing 21.61 sec
- by AR 6.97 sec

Rework (Errors) **Quality 0.08x**

- by Drawing 107 12.59%
- by AR 9 1.06%

The multipliers on the right are factors by which one can multiply your current value-added process costs for labor and rework to approximate the costs of those same processes in an AR environment.

Why do this?

The first question which is quite reasonably asked is "why do this as a game – why not use a real production aircraft assembly"? The most direct answer to this question is that NDA's prevent the publication of the known success stories of AR implementation. On the infrequent occasions when some information is released for public dissemination, it rarely (actually *never*) has sufficient context to make an evaluation of how that might apply to another situation. Here are some examples:

"…saving 300 hours just on grip measurements alone…"

"…time savings per shipset … 111 hours."

"During our first year of use, we had zero rework in the area where we use AssemblyWorks. A cost reduction of over $10M compared to the prior year."

"Our bracket installation is now only 34% of the previous assembly method…"

"…a cumulative recurring flyaway-cost savings of more than $82M."

"Fastener installation was reduced by 150 hours on the mid-fuselage…"

"…96 hours saved per ship set for the wing assembly…"

"AssemblyWorks reduced our nacelle assembly by 80 hours…"

"…layout labor costs have been reduced by approximately 50%, and rework is down by 99%."

"On our very first use, a 14 hour task was done in 4 hours with zero errors."

"…we just completed our first panel using AssemblyWorks. It was the first one we ever made with no rework."

When reading the quotes above, you may think to yourself that it seems very likely that all of those statements are reporting good news. But, you do not think – "well, that obviously applies to me – it will clearly make my process better too". So, the purpose of these time trials is to have a laboratory-quality controlled environment where we can compare a large sampling of people performing an identical task twice – once with AR and once without. Their performance is timed in both cases and errors evaluated. The difference between the two is a measureable amount of *Human Performance Enhancement* generated by the *Augmented Reality* technology. Significantly, we have no limitations on supplying all of the details necessary to provide a context for the statistical information provided so that the reader can determine to what extent these values apply to their own situation.

Of the 85 participants in this experiment, they identified themselves as the following:

Manager	21.4%
Engineer	55.4%
Technician	12.5%
Other	10.7%

The Experiment

Data was collected over two days of sampling the attendees of the AeroDef conference in Long Beach, California on 9-10 February 2016. The first and most important thing to understand about the experiment is that we are always evaluating the same person doing the same task two ways. It is not a case where maybe all the fast people did the AR method, and the slow people did the traditional method. All 85 of them did both methods. Here are the key things to know about the experiment (the details are described later in the paper, for those who want to dig into them):

First of all – THE MOTIVATION: the motivating factor for participating (besides the competitive fun of it) was the chance to win an iPad4 for the fastest player (one awarded each day). The participant's score is the total time for the two trials added together plus a 10 second penalty for each error (this is the "charge" for rework). Thus, each player is motivated to perform their best in both cases, while being very careful not to make an error in either case. (One error would leave no chance to win.) We thought a $400 prize for a fun way for people to spend about 5 minutes of their time would provide a good motivation for people to try their best – and that seemed to prove true.

There were 10 variations of the assembly. These are very similar, but not perfectly identical. This is to prevent a person from watching a previous participant and having an advantage. Also, it minimizes any advantage of which method is performed second. There were 8 possible "part numbers", of which only four were ever used in one assembly. There are 153 possible insertion points for the parts, but each variant had only 10 parts to install. There were quantities of 4, 3, 2, and 1 of the four part numbers used. In all of the quantity-3 cases, there were always 2 adjacent placements.

Prior to starting, the participant would roll a pair of dice which would determine three things:

- Which method would be built first (traditional drawing or AR)
- Which variant would be built via traditional methods
- Which variant would be built via AR

So, for each participant, of the 10 variant possibilities, it is purely by random chance what variant will be built by what method, and which will be first. In an informal exit poll of about 20% of the participants, there was unanimous agreement that the trials could not bias the outcome in favor of AR. This was, of course, by design. If anyone felt like the trial was biased at all or did not represent "the real world", it would have the opposite of the intended effect.

The numbers in these charts and graphs are for "value-added" work only. Even a system that can assure 100% efficiency produces no yield at all when the resource is not being applied to the assembly. All of the time spent by a person walking from the assembly to the parts bin, selecting the right parts, and walking back to the assembly accomplished putting exactly zero parts into the assembly – thus added no value to the assembly during that entire time (we all love those spaghetti charts!).

This paper is limited to a discussion of value-added work specifically. The experiment was carefully crafted to minimize non-value-added work. It should be noted that *AR is an enabling technology* for systems that can eliminate non-valued-added work from the production line in some cases, or significantly reduce it in other cases. That is however, a topic for another paper.

So, as you are considering the implications of the seemingly impossible statistics described in this paper, bear in mind that the measurements encompass the value-added part of the work exclusively. How the multipliers apply your situation must be factored with a ratio of value-added-time/total-time for your process.

The Stats

As noted above, a dice roll determines which method would be used first. The results were:

# of Trials where Drawing was FIRST:	45
# of Trials where AR was FIRST:	40
Total Trials:	**85**

The chart below demonstrates how using either method first affects the build-times. Both methods have a delta equal to about 12% of the standard deviation, indicating that the difference is "in the noise" of the statistics. The fact that the AR time actually increases when executed second, which is counter-intuitive, provides another indicator that the difference is statistically insignificant. Being first or second does not bias the data in an appreciable way.

	FIRST	SECOND	Delta	
Avg Drawing TIME:	81.75	79.03	2.73	seconds
Avg AR TIME:	23.23	24.17	-0.95	seconds

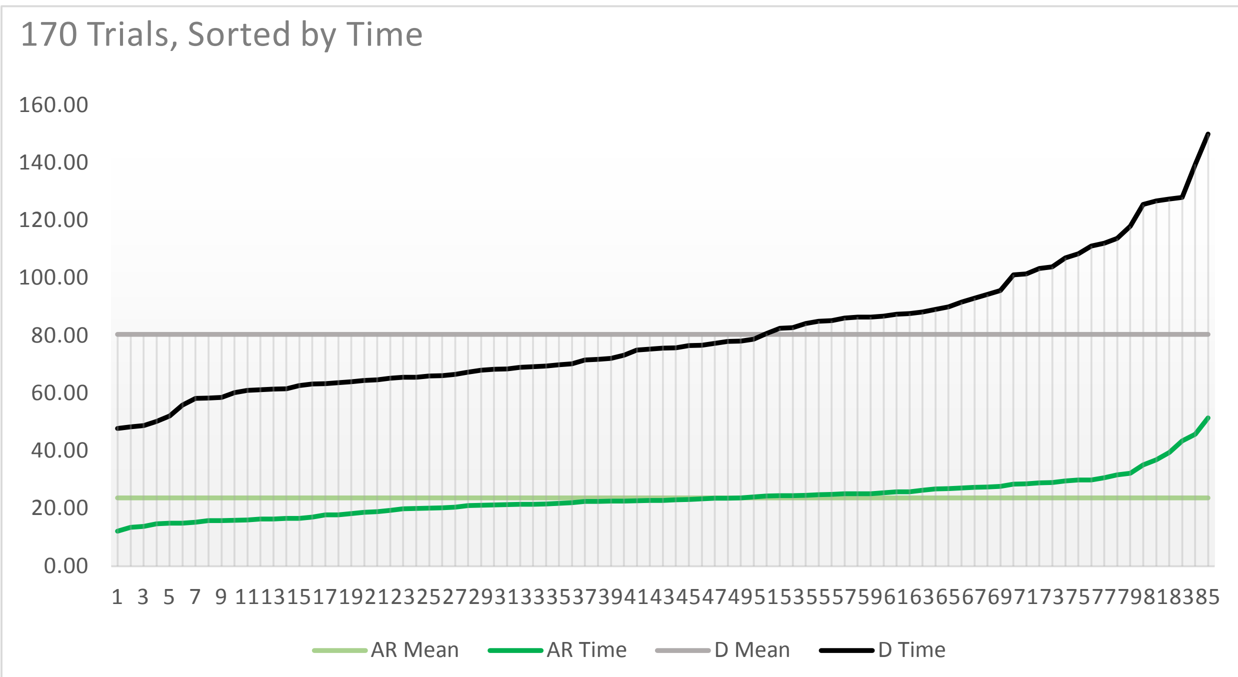
This next chart is not what we expected (hence the delta values are show as negatives). It would seem logical that after you played the game once, there would be a slight advantage for the second play. That was not the case. There were significantly more errors during the second try than the first for both AR and the drawing. Perhaps it is because the sample size of 85 is too small to be significant. If this is the result of fatigue; this is after less than 1.5 minutes of effort when drawings went first, and less than half a minute of effort when AR went first. Perhaps it was due to "pressure" to rush the second build from the feeling that the first effort was not enough. Until we have enough data to unravel this mystery, we will call this another statistically insignificant factor.

	FIRST	SECOND	Delta
Total Drawing ERRORS:	46	61	-15
Total AR ERRORS:	2	7	-5

The *9 AR errors* were made by 4 of the 85 participants. In all four cases the error was the same – the participant selected the incorrect part from the parts bin, and placed it in the correct position. The

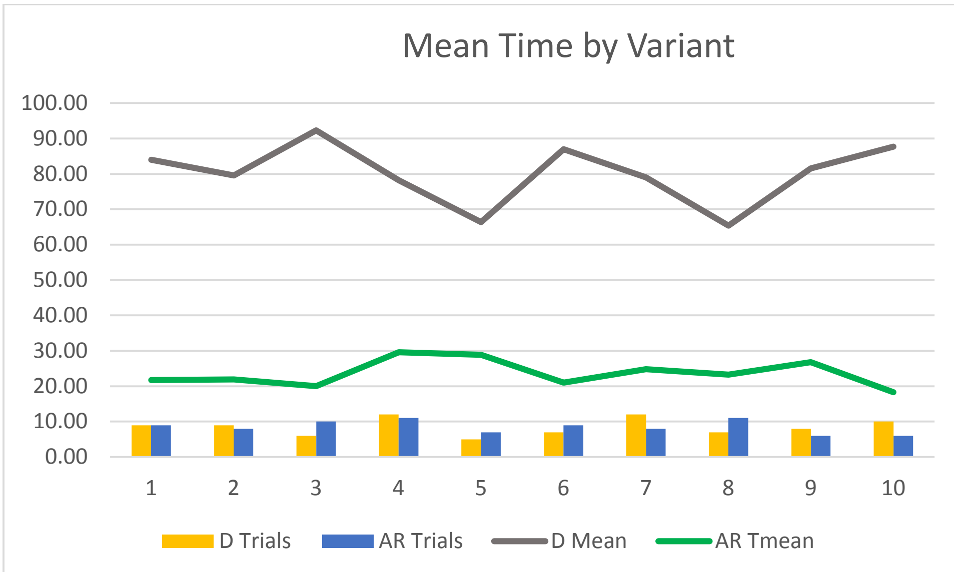
107 drawing errors were made by 23 people – over a quarter of the participants made at least one error when building from a drawing. These errors were roughly equally composed of all three error types: the correct position was left empty, an incorrect position was filled, or a correct position was filled with an incorrect part number.

The graph below shows the times of the 85 participants (sorted by trial times). The flatness of the AR graph (AR Time) relative to the drawing graph (D Time) indicates that there is much less variability in build-time associated with "who" builds the assembly. As noted earlier, the standard deviation is about one third for AR what it is when building from a traditional drawing. Load balancing a complex assembly line is made much simpler when the work force can move between work stations with absolutely no loss of capability. The knowledge of how to build the assembly should be stored in the AR system – not the heads of the technicians. That knowledge is shown to the technician at exactly the moment it needs to be known. Using this technique, every technician has the exact same knowledge (theoretically, the perfect knowledge) – making everyone more valuable.



We did not collect age data on participants, however we can say that the graph above which shows the times of the 85 participants (sorted by trial times) would track pretty close with age. Age 20 being the left side and 60-something being the right side of the chart. Note the sharp knee in the drawing curve (D Time) at the left side. The 20-somethings had the 6 fastest times, which were sharply better than others. However, they were not able to attain such markedly faster times using *AR* (only slightly faster). The winners both days were college sophomores.

Although there is nothing statistically relevant here, some people wanted to know if the 10 variants were really of equal difficulty. This graph shows the 10 variants across the page. The bar graph along the bottom shows how many times each variant was built by which method (selection was by a roll of the dice). If one variant was particularly more difficult or easier than another, we would expect to see the two lines have a correlation (both being fast or slow) at that point. That is never the case. For example, variant 5 was one of the fastest for drawings, but the slowest for AR while variant 10 is the opposite. Variants 3 and 8 are peaks on the drawing graph, but near the middle for the AR graph.



The Setup for the Experiment

Several items are attached as appendices to this paper. These are documents that were posted and shown to participants of the experiment.

- Appendix A: The rule sheet was shown to each player prior to playing
- Appendix B: A posted sign that provided additional rule explanations for those who were interested

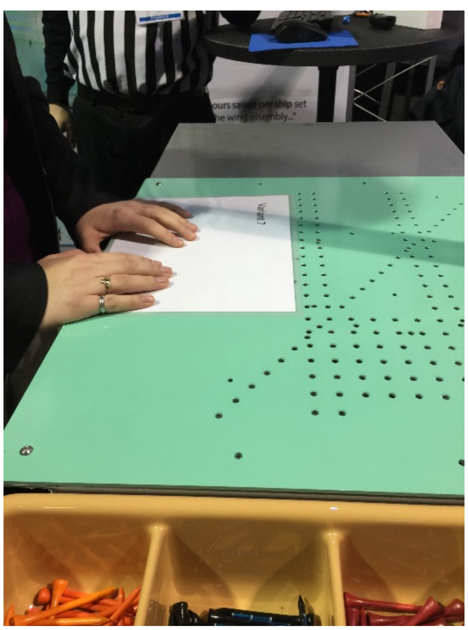
Execution of the Experiment

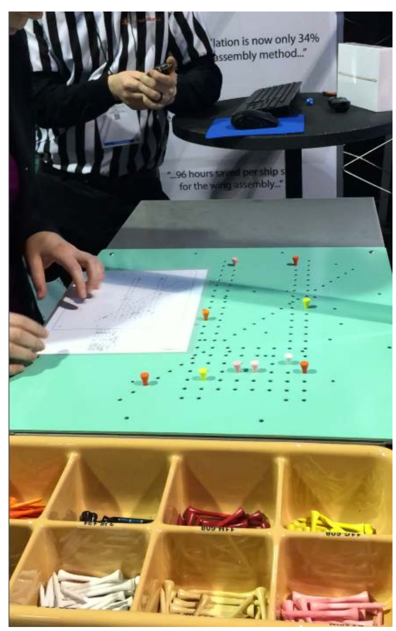
When a participant walked up and indicated they wanted to play, the "referee" would go over the rule sheet (appendix A) and explain the objectives, how to play, and how the game is scored. The player would then roll a d20 and a d10 dice. If the d20 value came up 11-20, then AR would go first and would use the variant number in the one's place of the d20; the drawing would go second and the variant was selected by the value of the d10. If the d20 value was 1-10, then the drawing would go first and would use the variant number of the d20; the AR method would go second and the variant was selected by the value of the d10. So, which goes first and what variant is built by what method is randomized. The chart above indicates that all the potential outcomes came out roughly equal.

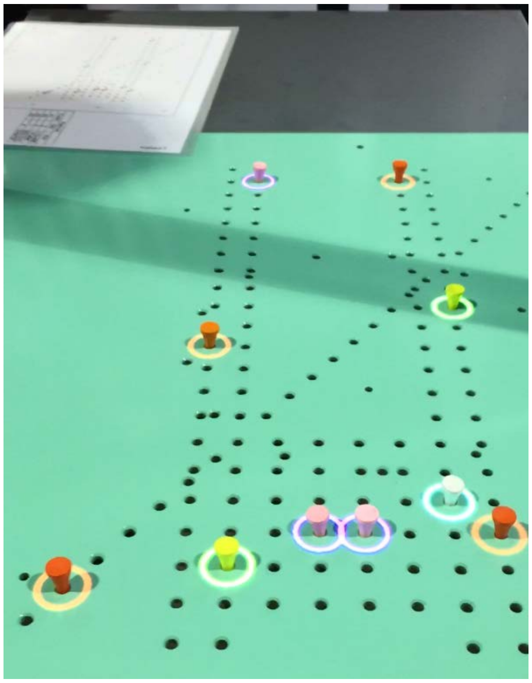
Now we know which variants are going to be built by which method and which goes first. The referee would go over the specifics of how the timing will work, and provide an opportunity for the player to familiarize themselves with the parts bin and the overall configuration. The stopwatch was started on either the start of projection, or the turning over of the drawing. The stopwatch was stopped when the player called “done”.

Colored pegs (golf tees) were used as our “fasteners” because it is fast and easy and accurate to check for errors.

When the player is building from the drawing they will turn the drawing over to start the clock (below left). Pegs are identified as part numbers in the parts bin. The player places the “fasteners” in the correct positions using the drawing in the same manner that fasteners are traditionally placed in the correct position of a complex assembly, such as an aircraft.

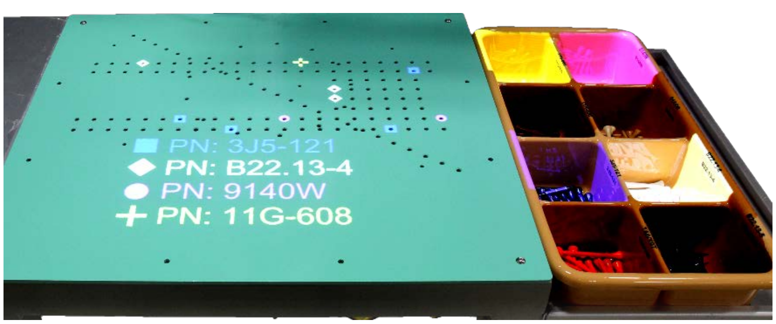






At the conclusion of the build of either type, the AR system is used to assist the player and referee to jointly inspect the assembly for errors (left). In this case there are no errors, however as can be readily seen, errors are very noticeable when they occur.

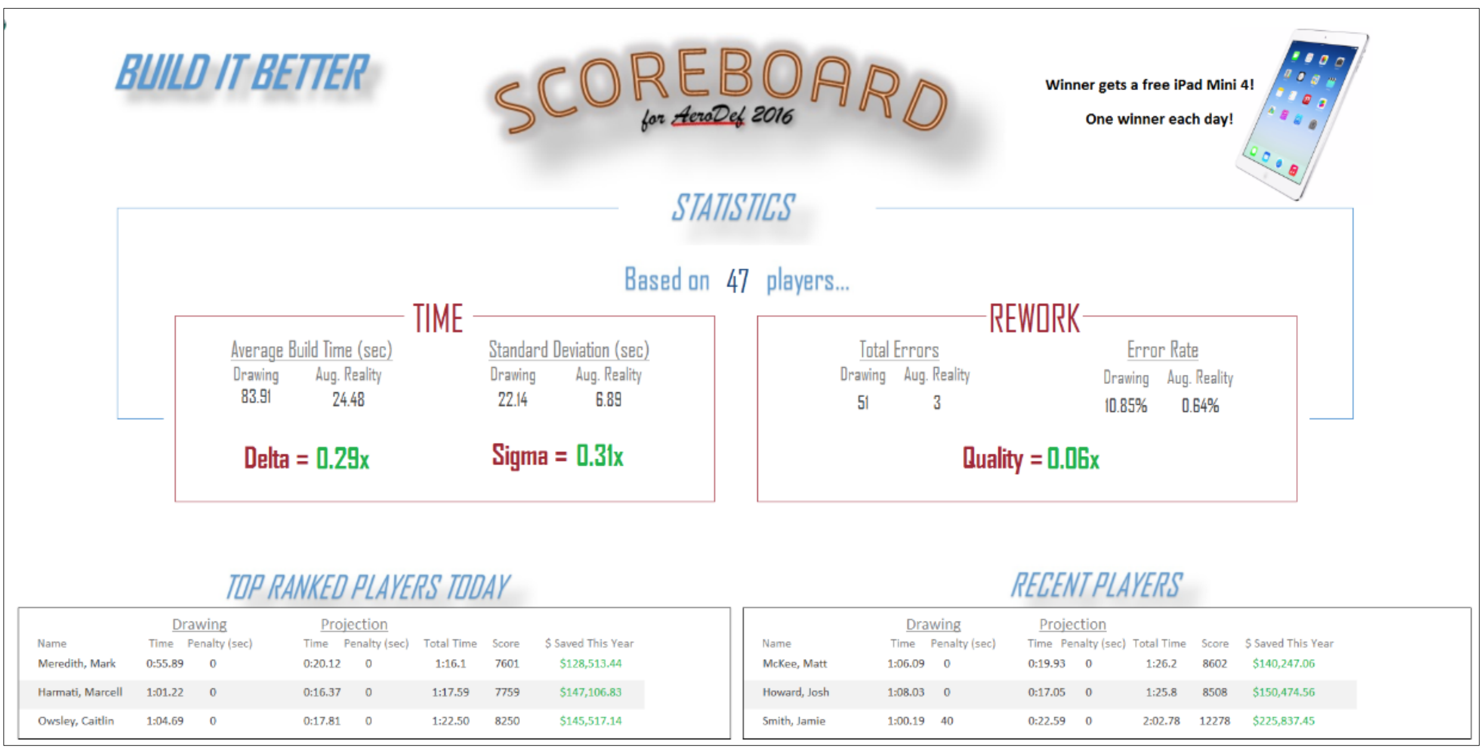
When building from AR (below), the participant was allowed to choose between a “show all” which shows all 10 parts using colors and shapes to distinguish the 4 part numbers; or a “sorted” display which shows the part number and quantity and uses a remote to step through the 4 part numbers one-at-a-time. In the sorted display, the presentation is always in white and the single parts bin is illuminated to indicate the correct part number. In the show-all display, the four parts bins are illuminated all at once with matching colors, shapes, and text.



Show All

Sorted

Information was captured in an Access database, from which the statistical data for this paper was obtained. Access also ran the scoreboard which is shown below. The scoreboard generated a lot of interest in the experiment. The dollar values shown are from a simple calculation using a $100/hour fully burdened rate, and $100 per rework. The dollars saved are the cost difference between the trials, annualized.



TOP RANKED PLAYERS TODAY

	Drawing		Projection				
Name	Time	Penalty (sec)	Time	Penalty (sec)	Total Time	Score	$ Saved This Year
Meredith, Mark	0:55.89	0	0:20.12	0	1:16.1	7601	$128,513.44
Harmati, Marcell	1:01.22	0	0:16.37	0	1:17.59	7759	$147,106.83
Owsley, Caitlin	1:04.69	0	0:17.81	0	1:22.50	8250	$145,517.14

RECENT PLAYERS

	Drawing		Projection				
Name	Time	Penalty (sec)	Time	Penalty (sec)	Total Time	Score	$ Saved This Year
McKee, Matt	1:06.09	0	0:19.93	0	1:26.2	8602	$140,247.06
Howard, Josh	1:08.03	0	0:17.05	0	1:25.8	8508	$150,474.56
Smith, Jamie	1:00.19	40	0:22.59	0	2:02.78	12278	$225,837.45

The following pages include the information that was provided to the participants prior to their time trials.

AssemblyWorks: The Game

Faster Builds. Less Rework. Proven Results.

OBJECTIVE: Install 10 “fasteners” into a small panel using two methods:

Traditional Paper Drawing Augmented Reality

There are 10 variations of the game

Each variant uses 10 fasteners (colored pegs) of 4 random colors:

- 4 of color A
- 3 of color B
- 2 of color C
- 1 of color D

How To Play

- The player will complete 2 stopwatch time trials: a paper drawing trial and an augmented reality trial
- The player rolls a d20 and d10 die to determine which game variants they will play and which one is first
- The player chooses “Show All” or “Sorted” for their Augmented Reality trial (see reverse side)
- Play starts with all pegs in the storage trays
- On the start signal, the player will place the correct color pegs in the correct holes. Limit 2.5 minutes per trial.
- When finished, the player will immediately notify the game host and the game host will stop the stopwatch
- The game host records the trial time, and the player then begins the 2nd trial

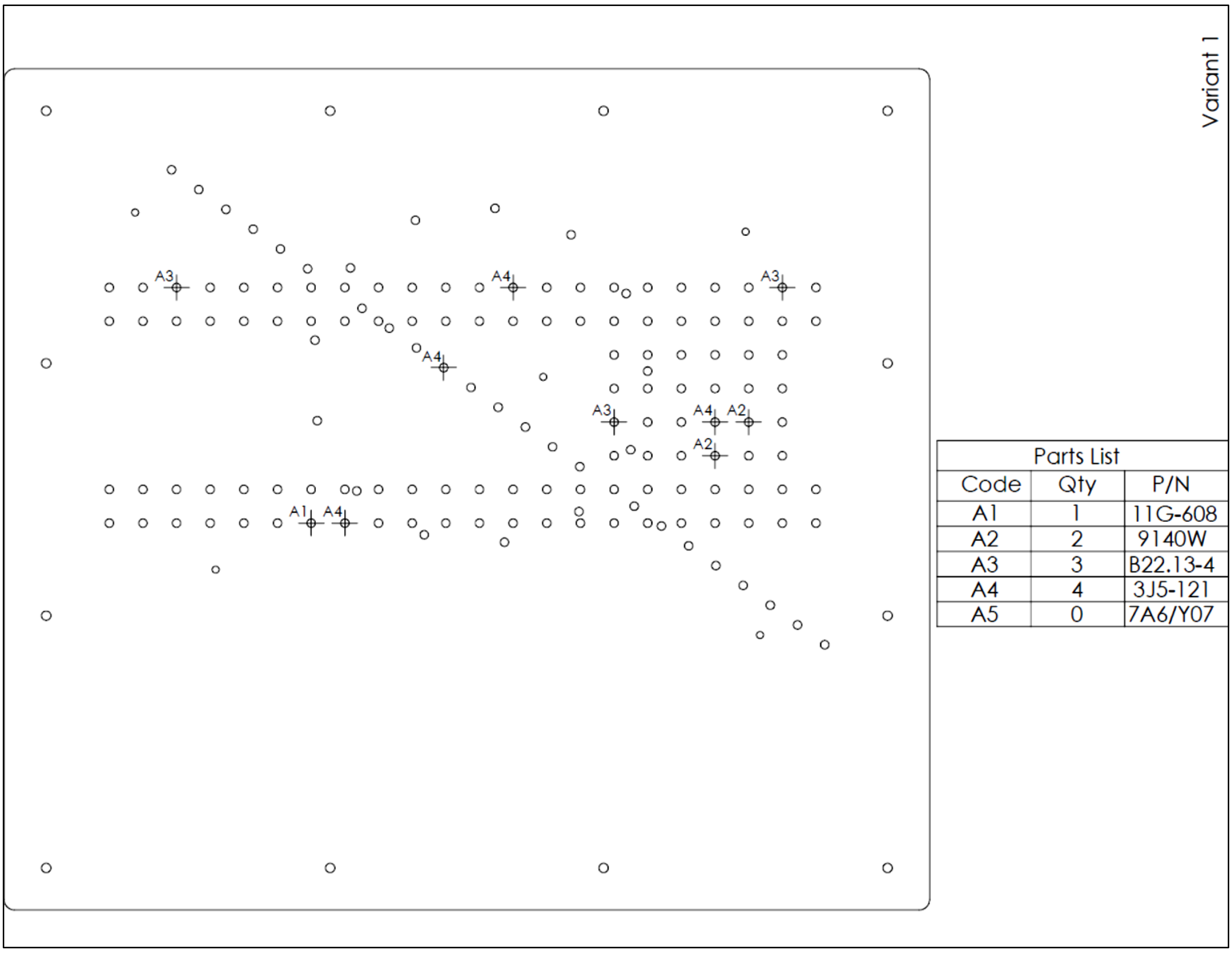
SCORING

- Errors require rework to correct. In this game, there is a 10 second penalty per error:
  - Any Peg installed in unmarked hole = 1 error
  - Peg not installed in marked hole = 1 error
  - Wrong Peg in marked hole = 2 errors
- YOUR SCORE: the sum of the two time trials and rework penalties
  - Player must have at least 8 correct pegs in both trials to be eligible to win

THE LOWEST SCORE WINS.

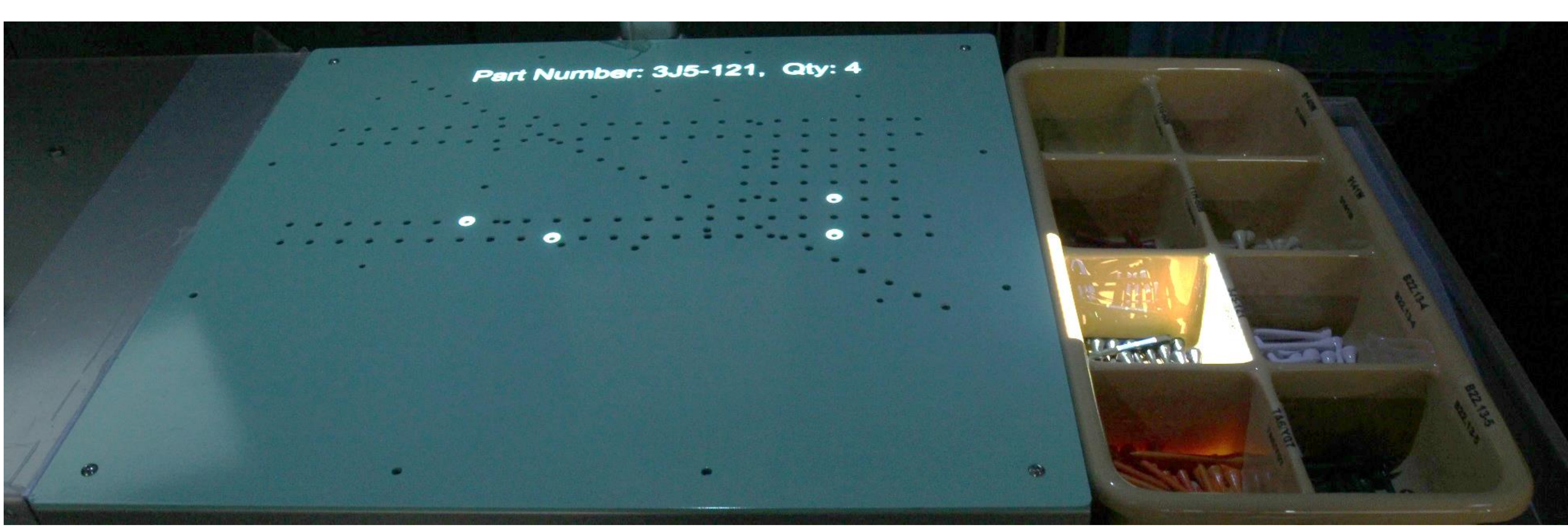
ONE PLAY PER PERSON PER DAY, PLEASE.

Expanded rules posted in back of booth



Drawing Sample

Augmented Reality Show <u>All</u> – All 10 fasteners are shown at once using unique colors and shapes. All 4 bins are illuminated in respective colors. A key is projected in the respective color naming the part number and showing the shape used.



Augmented Reality <u>Sorted</u> – The 10 fasteners are shown in 4 steps, 1 step per part number. One bin is illuminated while the locations of the bin's contents are projected in white. All projections of this type are white in this game as colors are not necessary. The respective part number and quantity is projected at the top of the panel. The player will be given a remote to step through the 4 steps. Upon pressing next, a new bin will be illuminated and the respective locations and quantities will be shown.

Dice, Scoring, and Game Variants

The player rolls a d20 and d10 die to determine which game variants they will play and which one is first

1. Two die are rolled – a d20 and a d10.
2. The value of the d20 determines which assembly method will be used first:
    a. **1-10** will be paper first
    b. **11-20** will be Augmented Reality first
3. The value of the d20 also determines which variant will be used for the first build. Examples:
    a. d20 = 7: Drawing variant #7 first
    b. d20 = 17: Augmented Reality variant #7 first
4. The value of the d10 will indicate the variant to be used for the other method.
    a. If it is the same as the d20, the d10 is rolled until values are different

Rework penalties and scoring

- Errors require rework to correct. In this game, there will be a fixed 10 second penalty per error
    - **Tee installed in unmarked hole = 1 error**
        - Rework required to remove wrong part
    - **Tee missing from the marked hole = 1 error**
        - Rework required to install correct part
    - **Wrong Tee in marked hole = 2 errors**
        - Rework required to remove wrong part
        - Rework required to install correct part

- YOUR SCORE: the **sum of the two time trials plus rework penalties**
    - Score is displayed as the above calculation in hundredths/second
    - Player must have at least 8 correct pegs in *each* of the two builds to be eligible to win
    - Player may play as many times as they want, but only the first trial in a day will count towards winning eligibility. Players that have not yet played have priority over repeat players.
    - Lowest score wins!
    - One winner each day wins an iPad Mini 4!

Game Variants

- The 10 variants all have "part numbers" (represented as colored tees) in different locations, but they all have the *same number of tees* and the *same number of different part numbers*.
- They are of equal difficulty. The variants are 10 similar patterns. Of the 8 potential part numbers – only 4 are ever used in one variant.
    - One part number (color) with a quantity of 4
    - One part number (color) with a quantity of 3
    - One part number (color) with a quantity of 2
    - One part number (color) with a quantity of 1
- Some variants may be temporarily suspended by the game host if people standing in line were in a position to observe them, giving them an unfair advantage over the previous player(s).

- Throughout the span of the event and across many players, the random selection method is intended to provide:
    - Equal number of builds with each method being first
    - Equal number of builds of each variant
    - Equal number of builds of each variant via each method

For any additional questions, please see the game host.